



Testing the Perceptual Relevance of Syntactic Completion and Melodic Configuration for Turn-Taking in Dutch

Johanneke Caspers

Phonetics Laboratory,
Universiteit Leiden Centre for Linguistics,
PO Box 9515, 2300 RA Leiden, the Netherlands
j.caspers@let.leidenuniv.nl

Abstract

The research presented in this paper focuses on the role of melodic configuration and syntactic completion in the turn-taking process in Dutch. Subjects were presented with fragments of task-oriented dialogue, in which syntactic completeness and four types of melodic configuration were systematically varied, asking them to indicate whether they expected the turn to change after the fragment or not. In a second test fragments containing speaker changes were presented, asking the subjects to indicate whether they thought the original speaker wanted to yield the turn or not. Results indicate that syntactic completion is the main factor in projecting possible turn-transition places: the number of expected speaker changes is very low when the current speaker has not reached a possible completion point. A rising pitch accent followed by a level boundary tone ($H^* \%$) is generally interpreted as a signal that the speaker wishes to continue, irrespective of syntactic completion, while $H^* H\%$, $H^* L L\%$ and $H^* L H\%$ configurations at syntactic boundaries are expected to be followed by a speaker change in the majority of cases. The data support the view that syntactic and melodic completion play a major role in the projection of possible turn-transition places.

1. Introduction

In everyday conversation there is generally a smooth and fast alternation of speaking turns, which can only be explained in terms of a highly complex system of interacting factors comprising syntax, semantics, pragmatics, prosody, visual cues etc. My specific interest is in the function of speech melody in the turn-taking process in Dutch.

In earlier research [1,2,3] a corpus of Dutch task-oriented (Map Task) dialogues was used to investigate the relationship between turn-taking, grammatical completion and local intonational markings. In the orthographic transcription of the materials projectable endpoints of utterances were indicated (cf. [4]), while the speech materials were divided into Inter Pausal Units (cf. [5]). The turn transition type of every IPU boundary was determined ('change' vs. 'hold') and a transcription of the melodic phenomena immediately preceding each boundary was made in the ToDI system [6]. Results showed that all regular changes of speaker occur at possible syntactic completion points (see also [7] for Dutch, [4] for English, [5] for Japanese). Speech melody supports syntax, in the sense that completion points that coincide with IPU boundaries are marked with a high or low boundary tone ($H\%$ or $L\%$) in 83% of the cases, while IPU boundaries

occurring within syntactic units are marked with an incomplete melodic configuration (ending in a level boundary tone, $\%$) in 71% of the cases. However, the rising pitch accent followed by level pitch ($H^* \%$) seems to function as an independent turn-keeping cue: it can be used to gap a syntactic break between two utterances produced by the same speaker (and there are no other combinations of pitch accent type and boundary tone type that behave the same).

The present investigation aims at testing the perceptual relevance of syntactic completion and a number of melodic configurations for turn-taking in Dutch. Are syntactic completion and melodic completion relevant for the perception of possible turn-transition places? Are they actual preconditions for turn-taking?

2. Approach

Fragments from the available Map Task materials (i.e., natural dialogues) were selected and presented to listeners in judgment tasks. In the first half of the experiment subjects were presented with only the initial part of each fragment, up to the position where a number of conditions was met (the 'target'), and then they had to indicate what they thought would happen immediately afterwards: the same speaker continues or the other speaker takes over. This way information about the projection of possible turn-transition places can be obtained. In the second half of the experiment the subjects were asked to focus on speaker changes occurring at a specific point and indicate what they thought the original speaker had intended: yield the turn or continue speaking. It did not seem fruitful to ask the subjects to judge the *acceptability* of speaker changes happening under certain conditions, since even overt interruptions may be perceived as perfectly acceptable in natural dialogue, especially when it is task-oriented.

2.1. Materials

The following variables were included in the design: plus versus minus turn change, plus versus minus syntactic completion and melodic type.

2.1.1. Turn Change

Half of the fragments contained a change of speaker ('change'), and the other half contained a pause longer than 100 ms after which the same speaker continued ('hold'). The fragments had a total duration between 8 and 15 seconds, depending on the amount of context that was necessary for the subjects to fully understand the interaction.



2.1.2. Syntactic Completion

The target occurred either at a syntactic completion point, or at a point where there was no syntactic completion. Note that care was taken that possible syntactic completion also meant possible pragmatic completion. Since speaker changes virtually always occurred at grammatical completion points in the Map Task materials, part of the data had to be generated artificially. To obtain speaker changes at non-completion points, fragments were chosen where a regular speaker change (taking place at a syntactic completion point) was preceded by a pause occurring at a non-completion point; then the stretch of speech between this pause and the change of speaker was removed from the fragment, resulting in a turn-change at a point where there is no grammatical completion.

2.1.3. Melodic Type

In addition to the rising pitch accent followed by a level boundary tone (H^* %), the same pitch accent type followed by a high boundary tone (H^* H%) was included in the design (the H^* pitch accent cannot be followed by a low boundary tone). In addition, the default pitch accent was included (the so-called 'pointed hat' or H^* L), followed by a low (L%) or a high (H%) boundary tone. This way the influence of pitch accent type (H^* vs. H^* L) as well as boundary tone type (% vs. H% and L%) could be investigated (to some extent, since the design was not complete). When possible, stimuli were selected where the relevant pitch accent and following boundary tone occurred on separate syllables. Furthermore, care was taken to spread the data over the eight Map Tasks and over the different speakers.

2.2. Method

For each combination of variables two examples were chosen, resulting in 32 basic stimuli. For all stimuli a fragment was created that was cut at the target (i.e., immediately after the boundary tone). Furthermore, from all stimuli containing a target followed by a change of speaker, a part was cut out containing only this speaker change, in order to indicate to the subjects on which particular speaker change to focus (which was necessary since many basic stimuli contained several speaker changes).

The data were presented to a group of 29 subjects, who were paid for their participation. In the first half of the experiment (part I) they listened to the fragments that were cut at the target position. Their task was to indicate on an answer sheet whether they thought the last speaker would continue speaking ('hold'), or whether the other speaker would take the following turn ('change'). Since for one of the four melodic types (viz., H^* %) 'hold' was expected to be the reply in the majority of cases, and since all targets at non-completion points were expected to lead to 'hold' anyhow, 'change' seemed a possible reply in only a third of the total number of cases. To avoid a skewed division of responses between two possible answer categories, a third category was therefore added: the backchannel [8,9]. This means that subjects could choose between the following expectations: (i) a change of speaker, (ii) the current listener produces an optional short background signal, and (iii) the current speaker continues. Since a backchannel is not taken as a

'real' speaker turn [6], the data can easily be reinterpreted as instances of either 'change' or 'hold' (i.e., (ii) plus (iii)). The fragments were presented twice.

In the second part of the experiment (part II) the subjects were presented with all basic stimuli containing target speaker changes ($N=16$); after presentation of the complete fragment, they heard only the part containing the relevant turn change, twice. They had to indicate on an answer sheet whether they thought the original speaker had expected the turn to change or whether (s)he had intended to continue; as a third possibility the category 'unclear' was used.

Both halves of the experiment started with three examples, after which the subjects could ask questions about the procedure.

It was hypothesized that the absence of syntactic completion as well as the absence of melodic completion (i.e., the presence of a H^* % contour) would lead to a majority of 'hold' responses in both parts of the experiment.

3. Results

3.1. Part I

The main results of the first part of the experiment are presented in table 1.

Table 1: Part I, absolute (and relative) scores per contour type, broken down by plus or minus syntactic completion.

contour	minus syntactic completion			total
	change	backchannel	hold	
H^* %	9 (8%)	-	107 (92%)	116
H^* H%	3 (2%)	75 (65%)	38 (33%)	116
H^* L L%	1 (1%)	7 (6%)	108 (93%)	116
H^* L H%	6 (5%)	71 (61%)	39 (34%)	116
total	19 (4%)	153 (33%)	292 (63%)	464
contour	plus syntactic completion			total
	change	backchannel	hold	
H^* %	11 (9%)	65 (56%)	40 (35%)	116
H^* H%	53 (46%)	51 (44%)	12 (10%)	116
H^* L L%	74 (64%)	36 (31%)	6 (5%)	116
H^* L H%	76 (66%)	29 (25%)	11 (9%)	116
total	214 (46%)	181 (39%)	69 (15%)	464

The table shows that the percentage of expected turn changes is very low for the points without syntactic completion (4%), irrespective of the preceding contour type, providing support for the hypothesis that syntax is the primary projection device in the turn-taking system. Backchannel responses are given in a third of the cases and they do present a clear effect of contour type: subjects expect them to follow a high boundary tone (for the H^* % contour the number of expected backchannels is even zero). In the remaining 63% of the cases a further turn from the same speaker is expected.

When the target coincides with a syntactic completion point, a clear influence of preceding contour type is visible on the number of anticipated turn changes. A rising pitch accent followed by a level boundary tone is expected to be followed by a change in only 9% of the cases, supporting the hypothesis that H^* % signals the wish of the current speaker to keep the turn (cf. [1,2,3,10]). For the H^* H% contour,



subjects have no clear expectations about what will happen afterwards (46% turn change, 54% hold). For both H*L pitch accents the other speaker is expected to take over in two-thirds of the cases, irrespective of the following boundary tone. Looking at the number of backchannel responses, the effect of the H% boundary tone seems to have vanished.

A hierarchical loglinear analysis performed on the factors 'score' (change, backchannel, hold), 'melodic type' and 'syntactic completion' reveals a significant association between 'melodic type' and 'score' (partial $\chi^2=232.1$, $p<.0001$) and between 'syntactic completion' and 'score' (partial $\chi^2=398.5$, $p<.0001$), and interaction between the three factors (Pearson $\chi^2=251.2$, $p<.0001$). Backward elimination leads to a model including all factors and all interactions. Results of partial χ^2 tests on the number of expected 'hold' scores for all combinations of contour types are presented in table 2.

Table 2: Part I, values of partial χ^2 tests (Pearson) on the 'hold'-scores for all pairs of contour types, broken down by syntactic completion (* indicates $p<.05$).

contour	minus syntactic completion		
	H* %	H* H%	H*L L%
H* H%	3.2		
H*L L%	6.7*	1.0	
H*L H%	0.6	1.0	3.7
contour	plus syntactic completion		
	H* %	H* H%	H*L L%
H* H%	38.1*		
H*L L%	73.7*	7.7*	
H*L H%	77.7*	9.2*	0.1

Table 2 shows that for the positions without syntactic completion there is one small – but significant – difference in the number of 'hold'-scores: between contours H* % and H*L L%, the two 'extremes' (92% vs. 99% 'hold'). For the stimuli ending in a possible completion point all pairs of contours differ significantly from each other, except for the two pointed hat contours. This means that both H* % and H* H% differ from all other contours in the number of expected turn continuations, while H*L L% and H*L H% do not differ from each other, suggesting that it is not just the boundary tone that is important, but that also the preceding pitch accent is relevant for the turn-taking system (cf. [1,2,3]).

3.2. Part II

Table 3 presents the data from the second part of the experiment. The data from one subject could not be used because of a misinterpretation of the instructions.

Table 3 shows that the percentage of 'change' responses again is low for the cases without syntactic completion, except for contour H* H%. Here the subjects indicate that they think the original speaker expected a change of turn in 30% of the cases, which may be explained by the fact that a rising pitch accent followed by a high boundary tone can be taken as the canonical question intonation. This would mean that the subjects (re)interpreted the syntactically incomplete utterance preceding the speaker change as a regular question. Since this was not the case in the first part of the experiment (where H* H% receives only 3% 'change' responses), this suggests that the actual change of speaker occurring in the

stimuli in part II influences the judgments of the subjects regarding the expectations of the speaker who loses the turn.

Table 3: Part II, Absolute (and relative) score per contour type, broken down by syntactic completion.

contour	minus syntactic completion			total
	change	unclear	hold	
H* %	4 (7%)	1 (2%)	51 (91%)	56
H* H%	17 (30%)	7 (13%)	32 (57%)	56
H*L L%	4 (7%)	1 (2%)	51 (91%)	56
H*L H%	7 (13%)	3 (5%)	46 (82%)	56
total	32 (14%)	12 (5%)	180 (81%)	224
contour	plus syntactic completion			total
	change	unclear	hold	
H* %	21 (38%)	9 (16%)	26 (46%)	56
H* H%	55 (98%)	-	1 (2%)	56
H*L L%	40 (71%)	10 (18%)	6 (11%)	56
H*L H%	52 (93%)	2 (4%)	2 (4%)	56
total	168 (75%)	21 (9%)	35 (16%)	224

The number of expected speaker changes is very high when the turn changes after a syntactically complete utterance, with the exception of the H* % contour. The rising pitch accent followed by a level tone leads to 38% 'change' judgments, which is low in comparison with the other three contour types, but higher than expected.

A hierarchical loglinear analysis performed on the factors 'score' (speaker expected change, unclear, speaker expected hold), 'contour type' and 'syntactic completion' reveals a significant association between 'contour type' and 'score' (partial $\chi^2=86.1$, $p<.0001$) and between 'syntactic completion' and 'score' (partial $\chi^2=38.1$, $p<.0001$), as well as a significant interaction between the three factors (Pearson $\chi^2=24.1$, $p<.001$). Backward elimination leads to a model including all factors and all interactions. Results of partial χ^2 tests for all combinations of contours are given in table 4.

Table 4: Part II, values of partial χ^2 tests (Pearson) on the 'hold'-scores for all pairs of contour types, broken down by syntactic completion (* indicates $p<.05$).

contour	minus syntactic completion		
	H* %	H* H%	H*L L%
H* H%	16.8*		
H*L L%	0.0	16.8*	
H*L H%	1.9	8.3*	1.9
contour	plus syntactic completion		
	H* %	H* H%	H*L L%
H* H%	30.5*		
H*L L%	17.5*	3.8	
H*L H%	27.4*	0.3	2.2

The data presented in the table show that for the turn changes at non-completion points, contour H* H% differs significantly from all other contours in the number of expected 'hold' scores. Furthermore, the H* % configuration is the only contour that differs from the other three in the number of expected 'hold' scores when there is a syntactic completion point.



4. Conclusions and discussion

In summary, the results of the first part of the experiment indicate that subjects rarely expect a change of speaker at a position where there is no syntactic completion, as was hypothesized. When a possible grammatical completion point is reached, the responses are clearly influenced by the preceding melodic configuration: after a rising pitch accent followed by a level boundary tone the expectation is that the same speaker will continue, while a change of speaker is expected in the majority of other cases. These results were expected and can be explained by the fact that the H* % contour does not end in a 'real' (i.e., low or high) boundary tone. This means that changes of turn are expected only at positions where syntax as well as prosody reaches a possible completion point. This concurs with the view that syntax is the primary projection device, and that local prosodic phenomena play a limiting role in the turn-taking system [3,4,5,7,11].

The data do not permit any conclusions about the separate influence of pitch accent type and boundary tone type in projecting a possible turn transition place. However, there does not seem to be a clear difference between the high and low boundary tones, as was reported earlier [1,2,3].

The unexpected influence of the H* H% contour on the responses to the grammatically incomplete stimuli in the second part of the experiment – 30% 'change' responses, as opposed to 3% 'change' responses in part I – may be explained in the following way. Subjects were presented with the complete (basic) stimulus first, after which they heard a smaller part of the same stimulus, containing only the two utterances before and after the target. This smaller part was presented twice. It seems that in a third of the cases involving a H* H% contour, subjects reinterpreted the syntactically incomplete initial stretch of speech as an elliptic and therefore complete utterance (which was possible because the relevant preceding context was not part of the presentation anymore) and then judged the turn change that followed as predictable. Without the following turn, and with the preceding context, virtually no one expected a change of speaker in these cases. This unforeseen by-product of the way the stimuli were generated underscores the strong tendency of the H* H% contour to be interpreted as signalling a question (in stead of signalling continuation, cf. [10]).

Also, the number of expected turn-changes following a H* % contour at a syntactic completion point were higher than expected in part II (38%, as opposed to 9% in part I), indicating that a change of speaker in these conditions – syntactic completion, but no melodic completion – is far from impossible. Inspection of the relevant data reveals that this relatively high number of 'change' scores is caused by one of the two relevant stimuli and may be explained by the fact that this stimulus contains a very long pause (1600 ms) between the target and the start of the turn of the other speaker (the mean duration of the post-target pause is 600 ms). The original speaker appears to be waiting for the other speaker to give some reaction (for instance, a backchannel), which causes the subjects to indicate that they think the original speaker expected to lose the turn. This effect could not occur in the first part of the experiment, since the stimuli were cut

off at the beginning of the pause. This finding shows that other, not systematically controlled, variables present in the data (in this case pause duration) may interact with the variables under investigation.

The results indicate that interlocutors indeed adhere to general principles of syntactic and melodic completion in the turn-taking process (cf. [4]). However, the present data suggest that at least melodic completion is not a genuine precondition for turn-taking, since it seems possible to overrule the turn-keeping effect of an incomplete melodic configuration with temporal factors.

Further investigations (in progress) will involve, among other things, manipulation of pitch contours. Subjects will be presented with stimuli in which (part of) the pitch contours are systematically varied, asking them to indicate the better fitting contour in a specific context.

5. Acknowledgments

This research was funded by the Netherlands Organization for Scientific Research (NWO), under project #355-75-002.

6. References

- [1] Caspers, J., "Pitch accents, boundary tones and turn-taking in Dutch Map Task dialogues", *Proceedings 6th International Conference on Spoken Language Processing*, Vol. I, 565-568, 2000.
- [2] Caspers, J., "Looking for melodic turn-holding configurations in Dutch", *Linguistics in the Netherlands 2000*, John Benjamins, Amsterdam, 45-55, 2000.
- [3] Caspers, J., "Local speech melody as limiting factor in the turn-taking system in Dutch" (in preparation).
- [4] Ford, C.E., and Thompson, S.A., "Interactional units in conversation: syntactic, intonational, and pragmatic resources for the management of turns", in E. Ochs, E.A. Schegloff and S.A. Thompson (eds), *Interaction and grammar*, Cambridge University Press, Cambridge, 134-184, 1996.
- [5] Koiso, H., Horiuchi, Y., Tutiya, S., Ichikawa, A., and Den, Y., "An analysis of turn-taking and backchannels based on prosodic and syntactic features in Japanese Map Task dialogs", *Language and Speech* 41, 295-321, 1998.
- [6] Gussenhoven, C., Rietveld, T., and Terken, J., "ToDI, Transcription of Dutch Intonation", interactive course, URL <http://lands.let.kun.nl/todi>, 1999.
- [7] Huiskes, M., "The role of the clause for turn-taking in Dutch conversations", Dissertation Utrecht University (in preparation).
- [8] Yngve, V.H., "On getting a word in edgewise", *Proceedings of the 6th Regional Meeting of the Chicago Linguistics Society*, 567-577, 1970.
- [9] Duncan, S., and Fiske, D.W., *Face-to-face interaction: Research, methods, and theory*, Lawrence Erlbaum Associates, Hillsdale, N.J., 1977.
- [10] Caspers, J. "Who's next? The melodic marking of question versus continuation in Dutch", *Language and Speech* 41, 375-398, 1998.
- [11] Selting, M., "On the interplay of syntax and prosody in the constitution of turn-constitutive units and turns in conversation", *Pragmatics* 6, 357-388, 1996.